The development and use of a decadal climate prediction system at GFDL

Anthony Rosati[1], Thomas L. Delworth[1], Shaoqing Zhang[1], Rich G. Gudgel[1], You-Soon Chang[1], Whit Anderson[1], Keith Dixon[1], Rym Msadek[1], William Stern[1], Gabriel Vecchi[1], Andrew Wittenberg[1], Xiaosong Yang[1], Fanrong Zeng[1], Rong Zhang[1]

1. Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

To be submitted to Journal of Climate

Corresponding Author:
Anthony Rosati
Geophysical Fluid Dynamics Laboratory/NOAA
201 Forrestal Road
Princeton, NJ 08540-6649
Email: tony.rosati@noaa.gov

Abstract

In this paper we describe the development and application of the GFDL decadal climate prediction system. This system consists of the GFDL CM2.1 global coupled climate model, an ensemble coupled assimilation system for producing initial conditions as well as a reanalysis, and estimates of past and future radiative forcing changes.

We assess the skill in this decadal prediction system through conducting and analyzing suites of prediction experiments initialized over the period 1961-2012. For each of the 52 years, there is a ten member ensemble of ten year experiments starting from observed initial conditions. Each of the hindcasts is also forced with time-varying estimates of radiative forcing changes. This experimental protocol is consistent with the near term prediction experiments of the Coupled Model Intercomparison Project Phase 5, and these experiments have been made available for assessment in the fifth Assessment Report of the Intergovernmental Panel for Climate Change.

The most notable predictable signal on decadal timescales is the warming trend in response to the external radiative forcing. However, a consistent predictive signal in the North Atlantic associated with Atlantic Multidecadal Variability emerges. The fact that this internal variability signal emerges over the external forcing and has predictive skill from the initialization is encouraging to the prospects of the nascent decadal prediction activity.

A decadal prediction starting in January 2012 and ending in 2021 is presented. Consistent with the uninitialized projections there is increased warming over the decade with greater warming over land. However, the initialized 2012 predictions are overall cooler than the uninitialized projections especially in the Southern Ocean.

1. Introduction

An important question in the study of climate change is whether ‘near term’ or decadal predictions of

45 future climate change could be improved if models begin their predictions from an observed based state
46 estimation of the climate system (e.g. Meehl et al. 2009; Meehl et al. 2012). The underlying issue is the
47 following: changes in the climate system are a combination of **internal variability** of the coupled
48 system and the response of the climate system to radiative forcing changes (**the forced response**).
49 Most previous climate change simulations have started from an arbitrary point in a long control
50 simulation, and then impose changing atmospheric composition. The detailed time evolution of the
51 model response is not expected to match the observed seasonal to decadal evolution of the climate
52 system, since the model's internal variability will not match that of the real climate system. This type
53 of simulation is only meant to estimate the forced response of the climate system. Given that both the
54 internal variability and the forced response are important sources of potential predictability in global
55 scale projections, it is important that the ensemble of predictions sample both model and initialization
56 uncertainties (Hawkins and Sutton, 2009). A comprehensive review of the mechanisms responsible for
57 internal decadal climate variability and the understanding of the dynamics of interdecadal variability
58 may be found in Liu (2012).
59 As part of the CMIP5 (Taylor et al 2012) and the IPCC AR5 assessment, the international community
60 is conducting a set of coordinated experiments in which models used for the prediction of climate
61 change are initialized with estimates of the observed state of the climate system. The key question is
62 whether this initialization process produces model simulations and predictions that are more skillful at
63 predicting the details of the future evolution of the climate system across time scales, from seasonal to
64 decadal and longer, than simulations that are not initialized. These simulations are designed to compute
65 both the forced response of the climate system and the time evolution of the internal variability of the
66 climate system starting from an observed state. Several studies have suggested that initialization could
67 improve skill (Smith et al. 2007, Keenlyside et al. 2008, Pohlmann et al 2009).
68 This near term forecast study uses the GFDL CM2.1 coupled model (Delworth et al 2006). The
69 CM2.1 model has been the subject of previous predictability work. Zhang (2008) showed that the

Atlantic Multi-decadal Oscillation (AMO) is tied to the North Atlantic subsurface temperature in observations and also in a long control run of CM2.1. Msadek et al (2010) using perfect model experiments demonstrated that the Atlantic meridional overturning circulation (AMOC) is potentially predictable on decadal time scales. Branstator et al. (2012) quantified the initial-value predictability of six coupled models for both the North Pacific and North Atlantic oceans. The predictability for the average 300m upper ocean temperatures from a CM2.1 1000-year control integration for the North Pacific was found to be of the order 9 years and for the North Atlantic of the order 12 years. These studies indicate that the CM2.1 model has substantial potential predictability, however, to what extent the aforementioned idealized potential predictability will translate to realizable predictions is the subject of this paper.

Throughout this paper the definition of the following terms are: forecast and prediction are used inter-changeably and denote the output of simulations that are initialized with an observation based state estimation; hindcasts are retrospective forecasts that are initialized with past observations and are not using observations that would not have been available at the initial time (although the radiative forcing includes volcanoes); and projections are an estimate of future climate states based on the forced climate response of a particular emission scenario.

This paper attempts to distinguish the role of the internal variability from the forced response in the GFDL decadal prediction system and is organized as follows:

In section 2, the model characteristics as well as the initialization procedure are described. In section 3, the results of the decadal predictions are assessed relative to observations before a discussion of the limitations to predictability in section 4. A prediction of global surface temperature out to 2021 is presented in section 5 and a summary and conclusions are given in section 6.

2. Initialization and Model

The feasibility of decadal predictions of internal variability largely stems from the role the ocean plays in the predictability of slowly evolving modes of variability. The challenge then is to have the capability to represent this low frequency climate variability within our climate models so that initializing them would offer the potential to predict internal variability in addition to externally forced climate change.

Initialization has three main components: the observing system, the assimilation method, and the model. These three components are combined to produce initial conditions for the climate model. We examine each component and its relation to the decadal prediction problem, focusing on the role of the ocean, as that is where the memory for predictability resides.

a. The Observing System

Historically, the sub-surface ocean has been very sparsely observed, and some of the data appear to be significantly biased (Domingues et al. 2008; Ishi and Kimoto, 2009; Willis et al. 2009), making the development and testing of ocean initialization schemes difficult. For instance, the non-stationary nature of the ocean observing system, particularly due to the paucity of salinity data as well as Expendable Bathythermograph temperature (XBT) data only going to 500-700 m depth, can give rise to spurious decadal variability making the assessment of forecasts difficult.

Studies of historical periods are important in order to assess the likely skill of forecasts over a range of different climate states. Recent and planned improvements to the observational network, however, offer significant potential for improvements in future forecast skill. Perhaps most important among these is the recent deployment of a global array of profiling floats by the Argo program (http://www.argo.ucsd.edu/). These provide, for the first time, contemporaneous measurements of both temperature and salinity over the upper 2 km of the global ocean (Roemmich and Gilson, 2009), potentially offering a step change in our ability to initialize ocean heat and density anomalies. These

measurements are, therefore, likely to be critical in order to make useful predictions of the Atlantic meridional overturning circulation and ocean heat transport (Dunstone and Smith, 2010, Zhang et al 2010). Another important recent contribution is the altimetry data (http://www.aviso.oceanobs.com) that, in addition to its own merits, holds great promise in conjunction with Argo.

b. The GFDL assimilation system

The GFDL assimilation system is based on the ensemble *adjustmen*t Kalman filter (EAKF; Anderson 2001), which is a deterministic variant of the ensemble Kalman filter. The EAKF estimates the probability distribution function (PDF) of climate states by combining the prior PDF derived from model dynamics and the observational PDF. It uses a two-step data assimilation procedure (the first step computes ensemble increments at an observation location and the second step distributes the increments over the impacted grids) for an ensemble Kalman filter under a local least squares framework. The filtering process is implemented by a multivariate linear regression with consideration of covariance of both atmospheric and oceanic state variables (Anderson 2003). The data adjusted ensemble members are the realizations of the analysis PDF and serve as the initial conditions for the next ensemble integration. The analysis steps are computed daily and distributed over the model time steps to reduce analysis shocks.

The GFDL system consists of an EAKF applied to GFDL's fully coupled climate model CM2.1 (Zhang et al. 2007), which is designed to produce a better-balanced initialization as opposed to each component model using its own assimilation system. The ocean component of the ensemble coupled data assimilation (ECDAv3.1) is the Modular Ocean Model version 4 (MOM4) configured with 50 vertical levels and 1° horizontal resolution, telescoping to 1/3° meridional spacing near the equator. The atmospheric component has a resolution of 2.5^{o} x 2^{o} with 24 vertical levels. A fully coupled model gives the first guess. The atmosphere is constrained by an existing atmospheric analysis. Ocean observations of temperature, salinity, and SST are assimilated using time-evolving covariance

142 structures from the coupled model. Argo observations are included as they became available in the
143 post-2000 period. The cross-interface covariance structures in the GFDL system allow for fully
144 coupled assimilation. For the ocean component, subsurface temperature and salinity observed profiles
145 (XBT, CTD, OSD, MBT, and MRB) from the World Ocean Database 2009 (Boyer et al. 2009) and
146 SST are assimilated daily (see details in Chang et al. 2012 and http://www.gfdl.noaa.gov/ocean-data-
147 assimilation). The atmosphere is constrained by an existing atmospheric analyses [NCEP/NCAR
148 reanalysis 1 for 1960-1978 (Kalnay et al. 1996) and NCEP/DOE reanalysis 2 for 1979-2010
149 (Kanamitsu et al. 2002)], using temperature, winds and sea level pressure. Sea ice and soil moisture
150 evolve without observational constraint.

151 All ECDAv3.1 experiments are performed with a 12-member ensemble that is used to compute state
152 estimation, ensemble mean, and the spread of the estimate. Ten of ensemble members are also used as
153 initial conditions. The ECDAv3.1 also uses covariance inflation that is designed to enhance the
154 consistency of upper and deep ocean adjustments, based on climatological standard deviation being
155 further informed by available observations (Zhang and Rosati 2010). The coupled methodology is
156 chosen to produce an initialization that is in a better-balanced state between the state variables of the
157 atmosphere and ocean. Ideally, the coupled assimilation would include the advantages of both the full
158 field initialization and anomaly initialization, in the sense that the initialization shock and model drift
159 issues are mitigated.

160 Chang et al. (2012) performed a comprehensive analysis of many aspects of the ECDAv3.1
161 reanalyses and showed that it is in good agreement with observations as well as existing reanalysis for
162 both climatology and variability. Throughout this paper the ECDAv3.1 reanalysis is used as our
163 verification source and will be referred to as **observations**.

164 ***c. Experimental Design***

165 We have used the GFDL CM2.1 climate model (Delworth et al., 2006) to conduct an extensive set of
166 hindcasts and predictions. The model is initialized using a state of the art coupled assimilation system

ECDAv3.1 (Chang et al., 2012, Zhang et al., 2007). For each set of observed states (initial conditions) we conduct a ten member ensemble of ten year hindcasts or predictions (hindcasts when the starting date is in the past, and so its accuracy may be assessed; predictions when the starting date is close to the present, and so its accuracy is not yet known). We use observed states from January 1 of each year from 1961 to 2012, for a total of 52 hindcasts and predictions (representing 5200 model simulated years). The ensemble members differ slightly in their initial conditions as derived from the assimilation system. The simulations use estimates of observed changes in radiative forcing that include volcanoes until 2005, and estimated forcing according to the Representative Concentration Pathways (RCP4.5) scenario (Meinshausen et al. 2011) thereafter. We refer to these simulations as the **initialized experiments**. It should be noted that the GFDL hindcast suite exceeds the CMIP5 recommendations (minimum of three ensemble members) for both number of ensemble members and number of start dates both of which should contribute to a better signal to noise ratio than would be achieved with fewer start dates and smaller ensembles.

In addition, we have conducted a separate 10-member ensemble of simulations using the CM2.1 model that are not initialized from an observed state. These start from arbitrary initial conditions in a control simulation, and cover the period 1861-2040. The simulations use the same changes in radiative forcing that the hindcasts use, and are intended to estimate the forced response of the climate system. We refer to these historical forcing simulations as the **uninitialized experiments.** The central question to ask is whether the initialized experiments provide better hindcasts than the uninitialized experiments.

d. Bias Correction

The CM2.1 climate simulation will drift away from the observed climate to the model's own climate as the forecast progresses beyond the initial conditions. In order to take the systematic error into account, we "bias correct" with a lead-dependent climatology. The protocol for full field initialization suggested for CMIP5 decadal forecasts (ICPO, 2011) was followed. The model drift is computed as the

192 difference between the ensemble of hindcasts and the observations averaged over the hindcast range,
193 specifically annual means from 1961-2011. The results presented here are validated to the
194 GFDL/ECDAv3.1 reanalysis and the drift from observations is relative to the ECDAv3.1 climatology.
195 The averages over the ensemble mean of the forecasts F and the corresponding observations O over
196 the forecast lead are:

197 $$\bar{F}_{nl} = \frac{1}{nfor}\sum_{nf=1}^{nfor} F_{nfnl} \qquad \bar{O}_{nl} = \frac{1}{nfor}\sum_{nf=1}^{nfor} O_{nfnl}$$

198 where nf = 1, . . . nfor is the number of forecasts, nfor=52 start dates from 1961-2012 and nl = 1, . . .
199 nleads is the number of annual mean forecast leads, where nleads =10 years. The model drift may be
200 defined as the difference between the ensemble mean of the forecasts and the observations averaged
201 over all start dates as:

202 $$d_{nl} = \bar{F}_{nl} - \bar{O}_{nl} \tag{1}$$

203

204 and the bias corrected forecast $\hat{F}_{nfnl}$ is obtained by subtracting the drift from each ensemble member
205 as:

206 $$\hat{F}_{nfnl} = F_{nfnl} - d_{nl} = \bar{O}_{nl} + (F_{nfnl} - \bar{F}_{nl}) = \bar{O}_{nl} + F'_{nfnl} \tag{2}$$

207 the anomaly of the forecast with respect to the forecast average is:

208 $$F'_{nfnl} = F_{nfnl} - \bar{F}_{nl} \tag{3}$$

209 and the observed anomaly is:

210 $$O'_{nfnl} = O_{nfnl} - \bar{O}_{nl} \tag{4}$$

211 The "bias correction" was calculated in a cross-validated manner where each corrected forecast does
212 not contribute to the forecast average, this is also done for the observations and unintialized projections

for consistency. Using start dates every year yields a bias correction that is systematic with lead and thus does not mask the predictable signal. Anomalies as defined by eqns. 3 and 4, will be used for all computations throughout this paper.

Figure 1 shows the zonal mean of the lead dependent drift ($\overline{d}$) for: air temperature at two meters (T2m); sea surface temperature (SST); and the average ocean temperature over the top 300m (Tav300), which is subtracted from the raw forecast to produce anomalies. The most prominent features in T2m are primarily a cold bias that gets progressively colder with lead and is at a maximum and more gradual at high latitudes, along with a warm bias in the Southern Ocean, which may be traced to a low-level cloud bias. The SST drift is nearly a reflection of the air temperature; however, Tav300 appears to delay the systematic error, which may be associated with predictability, as this would be the long lead memory in the system.

e. ***Hindcast Spread***

The use of ensemble standard deviation, as a measure of spread, is one approach to determine that the distribution of solutions of the individual ensemble member hindcasts show some reproducible behavior or whether it is chaotic. Figure 2 illustrates this for both SST and Tav300. At all time leads the SST has greater spread than the subsurface heat content, (except in the tropical West Pacific, due to winds/thermocline variations) which may be expected since SST would feel more of the effects of the internal variability of the atmosphere, whereas Tav300 would be subject to the greater thermal inertia of the ocean and thus contain the longer memory of the initialization to extend predictability. This shows that subsurface ocean, particularly in the Atlantic, is retaining some of the initial anomalies even for time leads of 5 and 10 years. This measure of irreproducibility or prediction uncertainty is not to be confused strictly with prediction error but rather potential predictability (or the model's capability of predicting itself (Boer, 2010)). There is little understanding on the relationship between spread and skill for climate predictions. Branstator et al. (2012) investigated the degree that the initial state can

influence climate predictions and it was found that for the North Atlantic in the GFDL model the memory of the initial state of heat content tends to persist for about a decade.

3. Decadal Prediction Results

The goal of this paper is to determine if, through initialization from an observed state there may be garnered additional predictability due to the representation of the internal variability over the uninitialized external radiatively forced signal. In order to discern this we have set up the best case comparison in that the coupled model and the forcing are identical in both the initialized and uninitialized experiments and the initialization and verification derives from a reanalysis that also uses the same coupled model and the same radiative forcing.

In order to assess skill, we have chosen to only show anomaly correlation coefficients (ACC) and for only three temperature variables. Although we also calculated mean squared skill score (MSSS), there did not seem to be any additional information content over the ACC.

Following Goddard *et al.* (2012) ACC is defined as:

$$ACC(nl) = \frac{\sum_{nf=1}^{nfor} (F'_{nfnl} * O'_{nfnl})}{\sqrt{\sum_{nf=1}^{nfor} F'^{2}_{nfnl} \sum_{nf=1}^{nfor} O'^{2}_{nfnl}}} \qquad (4)$$

where $ACC(nl)$ may be a particular lead or an average of leads.

The p-value from a two-tailed Student's t test is used for significance assessment of ACC. The p-value when taking the difference between correlation coefficients is from a two-tailed Fisher Z test for significance assessment (Goddard et al. 2012). The effective sampling size is used to reduce the degrees of freedom based on their autocorrelation (Bretherton et al. 1999).

Figures 3,4,5 show the global ACC at various forecast leads for variables T2m, SST, Tav300. The top row is the correlation between the uninitialized hindcasts and the ECDAv3.1 reanalysis (hereafter "observations"). The middle row is the correlation between the initialized hindcasts and observations. The bottom row is the difference between the two. Only positive values are assessed for significance, since positive ACC values indicate a positive linear relationship between the hindcasts and observations. As expected, the Year 1 initialized hindcasts show considerably more positive correlation that increases going from T2m to SST to Tav300, with the ENSO pattern being a dominant feature. The average of forecast Years 2-5 and Years 6-10 show very similar ACC patterns between the uninitialized and initialized hindcasts, indicating that the trend from the external forcing is dominant and causing both sets of hindcasts to converge. The difference maps show two exceptions: 1) Tav300 for Years 2-5 has more positive correlation years then T2m and SST. This indicates the more slowly evolving ocean heat content and longer predictability. 2) Every difference map shows persistent significant positive ACC in the North Atlantic region. To elucidate this feature, Figure 6 shows bias corrected hindcast SSTs where the top is the global average and the bottom is the average over the North Atlantic. For the global SST, the hindcast lead 1 is close to the observations, however, the leads 5 and 10 have converged to the uninitialized case, indicating that the forced climate response overwhelms the memory of the initial conditions. The natural aerosol forcing from the volcanoes El Chichon (1982) and Pinatubo (1991) is quite evident, although the response may be overly sensitive compared to observations (Delworth et al., 2005; Stenchikov et al, 2009).

In contrast the North Atlantic SSTs at all hindcast leads converge to observations (see also Pohlmann 2009) and follow the mean Atlantic multidecadal oscillation/variability (AMO/AMV) index (Yang et al, 2012) and the phase shift of the rapid warming starting in 1995. This feature is very distinct from the uninitialized ensemble-mean case where there is no rapid rise in temperature. What is particularly

interesting is that the internal variability manifested from the initial conditions clearly distinguishes itself from the uninitialized hindcasts in the North Atlantic.

In a companion paper, Yang et al. (2012) shows that the initialized GFDL decadal hindcasts may be capable of predicting SST Atlantic Multidecadal Variability up to 4 (10) year lead time at 95% (90%) significance level. Further Msadek et al. (in prep) investigates the predictability and the mechanisms of the North Atlantic warming within this GFDL decadal prediction framework. Within this paper we focus only on the question: was the 1995 North Atlantic warming predictable? To help to answer that question we show Figure 7, which depicts the North Atlantic Tav300 pentads starting in 1986 and ending in 2010. Looking at the observations (top row) the evolution of the increased heat content may be seen and the Year 1 lead (second row) shows good agreement. The Year 5 lead does not quite capture the rapid heat content increase in 1995 but once the warm regime is established the hindcasts are able to maintain it out to 2010. The Year 10 lead remarkably shows some skill also in the last two pentads. This is in contrast to Yeager et al. (2012) whose ensemble of decadal predictions successfully predicted the rapid rise in the mid-1990s heat content as early as 1991. Robson et al. (2012) found that the mid-1990s rapid North Atlantic warming could be understood as a delayed response to the prolonged positive phase of the NAO that led to a surge in northward heat transport. Clearly the top row of Fig. 7 shows that the ECDAv3.1 captures the evolution of the observed heat content anomalies and therefore our initialization is correctly preconditioned. To what extent the initial ocean memory of the heat content as well as the associated dynamic response gives rise to predictability is beyond the scope of this paper. Zhang et al. (2007) and Knight et al. (2005) suggest that the multidecadal variability in the Atlantic may potentially play some role in the evolution of Northern Hemisphere mean temperatures on decadal scales. Vecchi et al. (2012) shows that retrospective multi-year Atlantic hurricane predictions have high correlation in the initialized forecasts relative to the uninitialized projections. The increased skill of the initialized forecasts largely derives from the ability to persist the

initial conditions and thus represent the mid-1990's climate shift, which corresponded to the observed upward shift in Atlantic hurricane frequency. Although, as seen in figure 7 the predictions prior to 1995 did not successfully predict the shift. However, starting in 1995 and thereafter, the initialized forecasts were able to persist the climate shift and that is largely responsible for the increased skill. This is an excellent example, of the importance in considering the internal variability and the need to predict it.

Similar analyses were conducted for precipitation but the results are not shown, as there was no significant skill at any forecast lead or average of leads, except for forecast lead 1 in the tropics.

4. Limitations to Predictability

Decadal prediction is a cutting edge research topic and it is not clear what will be predictable. In order to assess the skill in decadal predictions the system first needs to be verified by hindcasts over a time-series of a long period (Kumar, 2009). Decadal prediction requires a long history in order to verify the low frequency inter-decadal variability. However, since the observational record is short and as in this study there are only 51 years, sampling uncertainty for decadal variability is to be expected. If the predictability were state dependent then the small sample size would make it difficult to determine in what state the prediction would be more reliable. The sampling issue is also compounded by the inhomogeneity of the observing systems, raising the question of the quality of the initial states. For example are the hindcasts of the '60s, with sparse XBT data and very little salinity data to constrain density anomalies, comparable to the hindcasts of the mid 2000s with Argo fully deployed globally? Yet in our verification all predictions are weighted equally. This validation is further exacerbated by the fact that the climate itself is not stationary. Of course, if the climate observing system is sustained and even supplemented, in time these sampling difficulties will be alleviated.

The relative importance of uncertainty amongst model, scenario, and internal variability is well represented in Figure 4 of Hawkins and Sutton, (2009). It shows that on decadal time scales the

dominant sources of uncertainty are model uncertainty and internal variability. In general the internal variability increases on smaller spatial scales and hence predictability is likely to be greater for large-scale ocean features and smaller for regional scale continental climate changes. Arguably both of these sources of uncertainty may be reduced as systematic errors in models are reduced and initialization is improved. Perhaps a possible rethinking of the role of scenario uncertainty on decadal time scales needs to be reconsidered given some of the recent findings on anthropogenic aerosols. As shown by Delworth and Dixon 2006, the aerosol forcing tends to cool the North Atlantic and enhance the AMOC. Many studies suggest that the observed multi-decadal variability in the North Atlantic may be linked to aerosol emissions (Rotstayn and Lohmann 2002; Even et al. 2009; Chang et al. 2012; Booth et al. 2012; Villarini and Vecchi 2012). This effect was not previously identified in IPCC/AR4 models since in most models only the direct effect of aerosols was represented; however, Booth et al. (2012) note that the inclusion of indirect aerosol effects has a large impact on North Atlantic SST. Although there is considerable uncertainty in the magnitude of aerosol forcing and its role in climate, it is likely to be an important consideration for decadal predictions.

The expectations for credible near term predictions are quite high among various stakeholders and probably too optimistic. The uncertainty in these experimental predictions needs to be conveyed as part of their delivery.

5. Temperature at 2m 2012-2021 Forecast

In the previous sections of the paper the hindcasts and their verification were the main topic. Here the decadal forecast initialized in January 2012 and using the RCP4.5 scenario as forcing projections is presented. Figure 8 shows the T2m anomalies relative to the 1971-2000 climatology. The first panel for the 2012 forecast shows the memory of the initial conditions as the cold eastern Pacific and warm North America features are prominent. The remaining three panels show the warming response from

the external forcing as well as the impacts of the initialization. The warming is stronger over land then ocean and over the Southern Ocean there is cooling that diminishes by the 2017-2021 forecast.

To try to put this in context Figure 9 shows the difference of the forecast anomalies with the uninitalized projections for pentad time periods of the decadal forecast. Overall the forecast is cooler then the uninitialized projections, however, regions of Asia, North Pacific and North Atlantic sub-polar gyre are warmer then the uninitialized projection. Consistent with Yang et al. (2012) the forecast shows the AMO to be still in the warm phase, resulting in the bi-polar T2m pattern. At the longer forecast leads the initialized forecasts approach the uninitialized projections.

6. Summary and Conclusions

Using the GFDL/CM2.1 coupled model and the ECDAv3.1 reanalysis, a series of initialized decadal hindcasts and forecasts as well as uninitialized historical runs were made to investigate decadal predictability. This experimental protocol is consistent with the near term prediction experiments of CMIP5 and IPCC AR5. Largely the decadal predictability is realized by the climate response to changing radiative forcing, particularly at forecast leads greater then three years. The North Atlantic emerges as a region where the internal variability, as represented by the initialization, showed skill out to long forecast leads that rose above that of the uninitialized projections. This is particularly evident in the rapid warming that occurred in the North Atlantic in 1995. This climate shift was not well predicted prior to 1995, however the predictions from 1995 and on were able to show skill for the shift.

Forecasts for the five-year periods 2012-2016 and 2017-2021 show that surface temperature will be warmer relative to the 1971-2000 climatology for all land regions and ocean basins, except for the Southern Ocean where cooling is predicted. The forecasts at leads greater then four years largely converge to the uninitialized projections, with the RCP4.5 scenario. Smith et al (2012) show the first multi-model decadal climate prediction, which includes the GFDL prediction amongst the ensemble.

This emerging effort will serve as experimental predictions that the entire climate community may be able to view and evaluate. Such an activity will enable comparisons of predictions and observations to identify model errors. This may well be the way forward toward advancing model skill.

Although formidable challenges remain, the potential utility of decadal predictions and real time attribution is enormous. While there is undoubtedly room for model improvement, we show that natural climate variability poses inherent limits to climate predictability. Thus raising expectations regarding predictability or the lack thereof should be avoided.

Acknowledgments:

The authors would like to express gratitude to Ronald Stouffer for his helpful review.

References:

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129**, 2884–2903.

Anderson, J.L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev*., **131**, 634-642.

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T. and Bellouin, N., 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature,* **484**, 228-232.

Boer, G.J., 2010: Decadal potential predictability of 21st century climate. *Climate Dyn.,* doi: 10.1007/s00382-010-0747-9

Boyer TP et al., 2009: World Ocean Data Base 2009, Chapter 1; introduction. In: Levitus, S (ed), NOAA Atlas NESDIS 66, U.S. Gov, Printing Office, Washington, DC, 216pp.

Branstator, G., H. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati, 2012: Systematic estimates of initial value decadal predictability for six AOGCMs. *J. Climate*, **25**, 1827-1846, DOI:10.1175/JCLI-D-11-00227.1.

Bretherton, C.S., M. Widmann, V.P. Dymnikov, J.M. Wallace, and I. Blade, 1999: The effective Number of Spacial Degrees of Freedom of a Time-Varying Field. *J. Climate*, **12,** 1990-2009.

Chang, C.-Y., J.C.H. Chiang, M.F. Wehner, A. Friedman, and R. Ruedy, 2011: Sulfate aerosol control of tropical Atlantic climate over the 20th century. *Journal of Climate*, **24**, 2540–2555.

Chang, Y.-S., S. Zhang, A. Rosati, T.L. Delworth, and W. F. Stern, 2012: An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation. *Clim. Dyn*., In Press, DOI:10.1007/s00382-012-1412-2.

Delworth, Thomas L., V Ramaswamy, and G Stenchikov, 2005: The impact of aerosols on simulated ocean temperature and heat content in the 20th century. *Geophysical Research Letters*, **32**, L24709, DOI:10.1029/2005GL024457.

Delworth, T.L., et al., 2006: GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and Simulation Characteristics. *Journal of Climate*, 19(5), DOI:10.1175/JCLI3629.1.

420 Delworth, T.L. and Dixon K.W., 2006: Have anthropogenic aerosols delayed a greenhouse gas induced
421 weakening of the North Atlantic thermohaline circulation? *Geophys. Res. Lett.,* **33,** L02606,
422 DOI:10.1029/2005GL024980.
423
424 Domingues, C.M., J.A. Church, N.J. White, P.J. Gleckler, S.E. Wijffels, P.M. Barker and J.R. Dunn
425 (2008), Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*,
426 **453**, 1090-1094, doi:10.1038/nature07080.
427
428 Dunstone, N. J. and D. M. Smith, 2010: Impact of atmosphere and sub-surface ocean data on decadal
429 climate prediction. *Geophys. Res. Lett*., **37**, L02709, doi:10.1029/2009GL041609,2010.
430
431 Evan, A.T., D.J. Vimont, A.K. Heidinger, J.P. Kossin, and R. Bennartz, 2009: The role of aerosols in
432 the evolution of tropical North Atlantic Ocean temperature anomalies. *Science*, **324**, 778–781.
433
434 Goddard, L., and Coauthors, 2012: A verification framework for interannual-to-decadal prediction
435 experiments. *Clim. Dyn.* (submitted).
436
437 Hawkins, E. and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions.
438 *Bull. Amer. Met. Soc.,* doi:10.1175/2009BAMS2607.1.
439
440 ICPO (International CLIVAR Project Office), 2011: Data and bias correction for decadal climate
441 predictions., International CLIVAR Project Office, CLIVAR Publication Series 150, 6pp.
442 (http://eprints.soton.ac.uk/171975/1/150_Bias_Correction.pdf).
443
444 Ishii, M., and M. Kimoto, 2009: Re-evaluation of historical ocean heat content variations with an XBT

depth bias corrections. *J. Phys. Oceanogr.,* **65**, 287-299.

Kalnay E et al (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull Am Met Soc 77:437-471.

Kanamitsu M, Ebitsuzaki W, Woolen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL (2002) NCEP-DOE AMIP-II reanalysis (R-2). Bull AmMet Soc 83:1631-1643.

Keenlyside, N., and Coauthors, 2008: Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, **453,** 84-88.

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M.E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.,* **32,** L20708, doi:10.1029/2005GL024233.

Kumar, A., 2009: Finite samples and uncertainty estimates for skill measures for seasonal predictions. *Mon. Wea. Rev.*, **137**, 2622-2631.

Liu, Z. 2012: Dynamics of Interdecadal Climate Variability: A Historical Perspective. *J. Climate*. DOI:10.1175/2011JCL3980.1.

Meehl, G. A., and co-authors, 2009: Decadal prediction, can it be skillful?, *Bull. Amer. Meteor. Soc.*, **90**, 1467-1485.

Meehl, G. A., and Coauthors, 2012: Decadal climate prediction: An update from the trenches. *BAMS* (submitted).

Meinshausen, M., S. J. Smith, K. V. Calvin, J. S. Daniel, M. L. T. Kainuma, J.-F. Lamarque, K. Matsumoto, S. A. Montzka, S. C. B. Raper, K. Riahi, A. M. Thomson, G. J. M. Velders and D. van Vuuren (2011). "The RCP Greenhouse Gas Concentrations and their Extension from 1765 to 2300." Climatic Change (Special Issue), DOI: 10.1007/s10584-011-0156-z.

Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett*., **37**, L19608, doi:10.1029/2010GL044517.

Msadek, R., A. Rosati, T. L. Delworth, W. Anderson, G. Vecchi, Y.-S. Chang, K. Dixon, R. Gudgel, B. Stern, A. Wittenberg, X. Yang, F. Zeng, R. Zhang, S. Zhang, 2012: Predicting North Atlantic decadal variability in the GFDL coupled system: the 1995 climate shift event, *in preparation.*

Pohlmann, H., Jungclaus, J., Kohl, A., Stammer, D. and Marotzke, J. 2009: Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic, *J. Climate,* 22, 3926-3938.

Robson, J., R. Sutton, K. Lohmann, D. Smith, and M.Palmer, 2012: Causes of the Rapid Warming of the North Atlantic Ocean in the 1990s. *J. Climate.* doi:10.1175/JCLI-D-11-00443.1, in press.

Roemmich, D. and J. Gilson, 2009: The 2004-2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo program, Prog. Oceanogr.,82, 81-100.

Rotstayn, L. D., U Lohmann, 2002: Tropical Rainfall Trends and the Indirect Aerosol Effect. *J.*

Climate, **15**, 2103–2116. doi: 10.1175/1520-0442.

Smith, D. M., and Coauthors, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796-799. Doi:10.1126/science.1139540.

Smith, D. M., and Coauthors, 2012: Real-time multi-model decadal climate predictions. Submitted Climate Dynamics.

Stenchikov, G, Thomas L Delworth, V Ramaswamy, Ronald J Stouffer, Andrew T Wittenberg, andFanrong Zeng, August 2009: Volcanic signals in oceans. *J. Geophys. Res.*, **114**, D16104, DOI:10.1029/2008JD011673.

Taylor, Karl E., Ronald J. Stouffer, Gerald A. Meehl, 2012: An overview of cmip5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498. doi: http://dx.doi.org/10.1175/BAMS-D-11-00094.1

Villarini, G., and G.A. Vecchi 2012: Twenty-first-century projections of North Atlantic tropical storms from CMIP5 models, *Nature Climate Change*, doi:10.1038/NCLIMATE1530.

Vecchi, G. A., and Coauthors, 2012,: Multi-year Prediction of North Atlantic Hurricane Frequency: Promise and Limitations. Submitted to *J. Climate.*

Willis J. K., J. M. Lyman, G. C. Johnson, and J. Gilson, 2009: In Situ Data Biases and Recent Ocean Heat Content Variability. *J. Atmos. Oceanic Technol.*, 26, 846-852.

519

520 Yang, Xiaosong, Anthony Rosati, Shaoqing Zhang, Thomas L. Delworth, Rich G. Gudgel, Rong
521 Zhang, Gabriel Vecchi, Whit Anderson, You-Soon Chang, Timothy DelSole, Keith Dixon,
522 Rym Msadek, William F. Stern, Andrew Wittenberg, Fanrong Zeng, 2012: A predictable
523 AMO-like pattern in GFDL's fully-coupled ensemble initialization and decadal forecasting
524 system. *J. Climate,* accepted.

525

526 Yeager, S., A., Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A Decadal Prediction Case
527 Study: Late 20th century North Atlantic Ocean Heat Content. *J. Climate*. doi:10.1175/JCLI-D-
528 11-00595.1, in press.

529

530 Zhang, R., T. L. Delworth, and I. M. Held, 2007: Can the Atlantic Ocean drive the observed
531 multidecadal variability in Northern Hemisphere mean temperature? *Geophys. Res. Lett*., **34**,
532 L02709, doi: 10.1029/2006GL028683.

533

534 Zhang, Rong, October 2008: Coherent surface-subsurface fingerprint of the Atlantic meridional
535 overturning circulation. *Geophys. Res. Lett.*, **35**, L20705, DOI:10.1029/2008GL035463.

536

537 Zhang, S. and A. Rosati, 2010: An inflated ensemble filter for ocean data assimilation with a biased
538 coupled GCM, *Mon. Wea. Rev*., **138**, 3905-3931.

539

540 Zhang, S., A. Rosati, and T. L. Delworth, 2010: The adequacy of observing systems in monitoring
541 AMOC and North Atlantic climate**.** *J. Climate*, **23**, 5311-5324.

542

543 Zhang, S., M.J. Harrison. A. Rosati, and A.T. Wittenberg, 2007: System Design and Evaluation of

Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Monthly Weather Review*, 135(10), DOI:10.1175/MWR3466.1.

548

549 ## Figure Captions:

550 **Figure 1:** Lead dependent zonal mean of the drift climatology for 2-meter air temperature (T2m), sea
551 surface temperature (SST), and averaged temperature over the upper 300 meters (Tav300). Contour
552 interval is 0.2 degrees C. Lead times are given in years.

553

554 **Figure 2:** Ensemble Standard Deviation of SST (left) and average ocean temperature over top 300
555 meters, Tav300 (right) from hindcasts for leads 1,5,10. Units are in degrees C.

556

557 **Figure 3:** T2m Anomaly Correlation Coefficient (ACC) with the GFDL reanalysis of the uninitialized
558 forecasts (top row), initialized forecasts (middle row), and their difference (bottom row). Forecast year
559 1 is the first column, the average of forecast years 2-5 is the second column and the average of forecast
560 years 6-10 the third column. The yellow contour lines represent the 90% significance.

561

562 **Figure 4:** Same as in Fig 3 but for SST.

563

564 **Figure 5:** Same as in Fig. 3 but for Tav300.

565

566 **Figure 6:** Two year running mean bias corrected average SST, black - OBS observations from
567 ECDAv3.1, red – Uninitialized , blue solid lead 1 forecast, blue dash lead 5, blue dot lead10. Top -
568 Global SST, Bottom – North Atlantic SST averaged over 30N-65N, 80w-10w. Units are in degrees C.

569

570 **Figure 7:** North Atlantic Tav300 pentads in degrees C. Top row shows observations. Subsequent rows
571 show forecast leads 1, 5, and 10. Leads are in years.

572

573 **Figure 8:** T2m forecast bias corrected anomaly relative to the 1971-2000 climatology. Units are in

574 degrees C.

575

576 **Figure 9:** Difference of initialized forecast (start date is January 1, 2012) and unitialized projection,

577 showing impact of initialization. Left – 2012-2016 forecast period, Right – 2016-2021 forecast period.

578 Contour units are in degrees C.

579

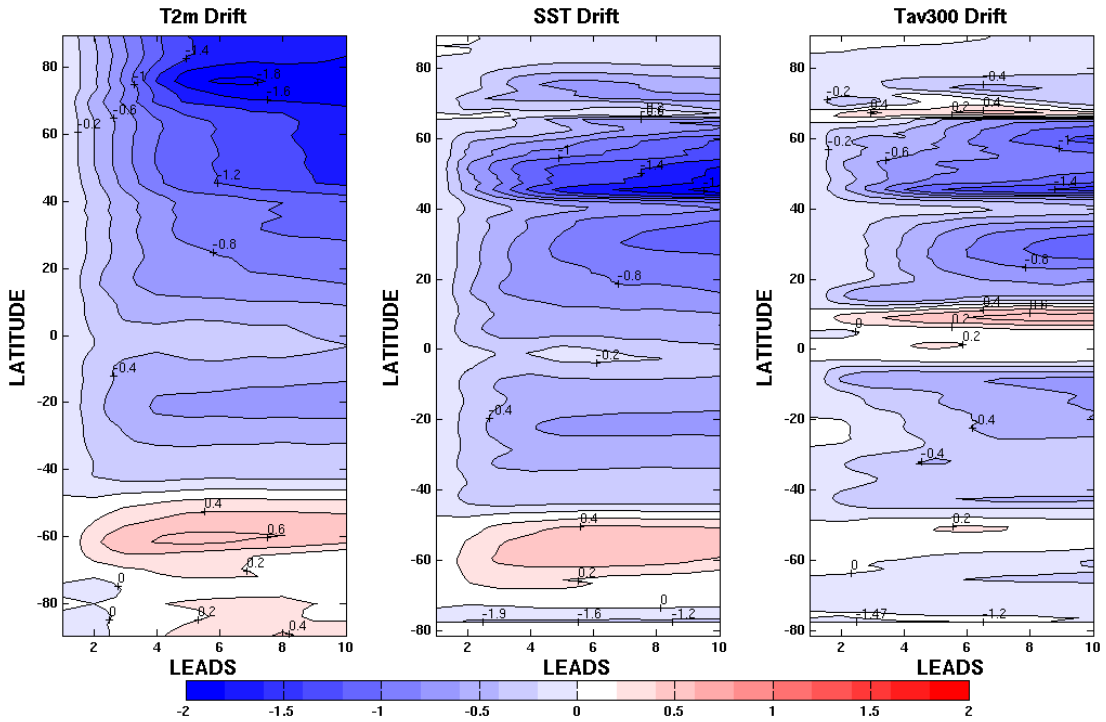


Figure 1 – Lead dependent zonal mean of the drift climatology for 2-meter air temperature (T2m), sea surface temperature (SST), and averaged temperature over the upper 300 meters (Tav300). Contour interval is 0.2 degrees C. Lead times are given in years.

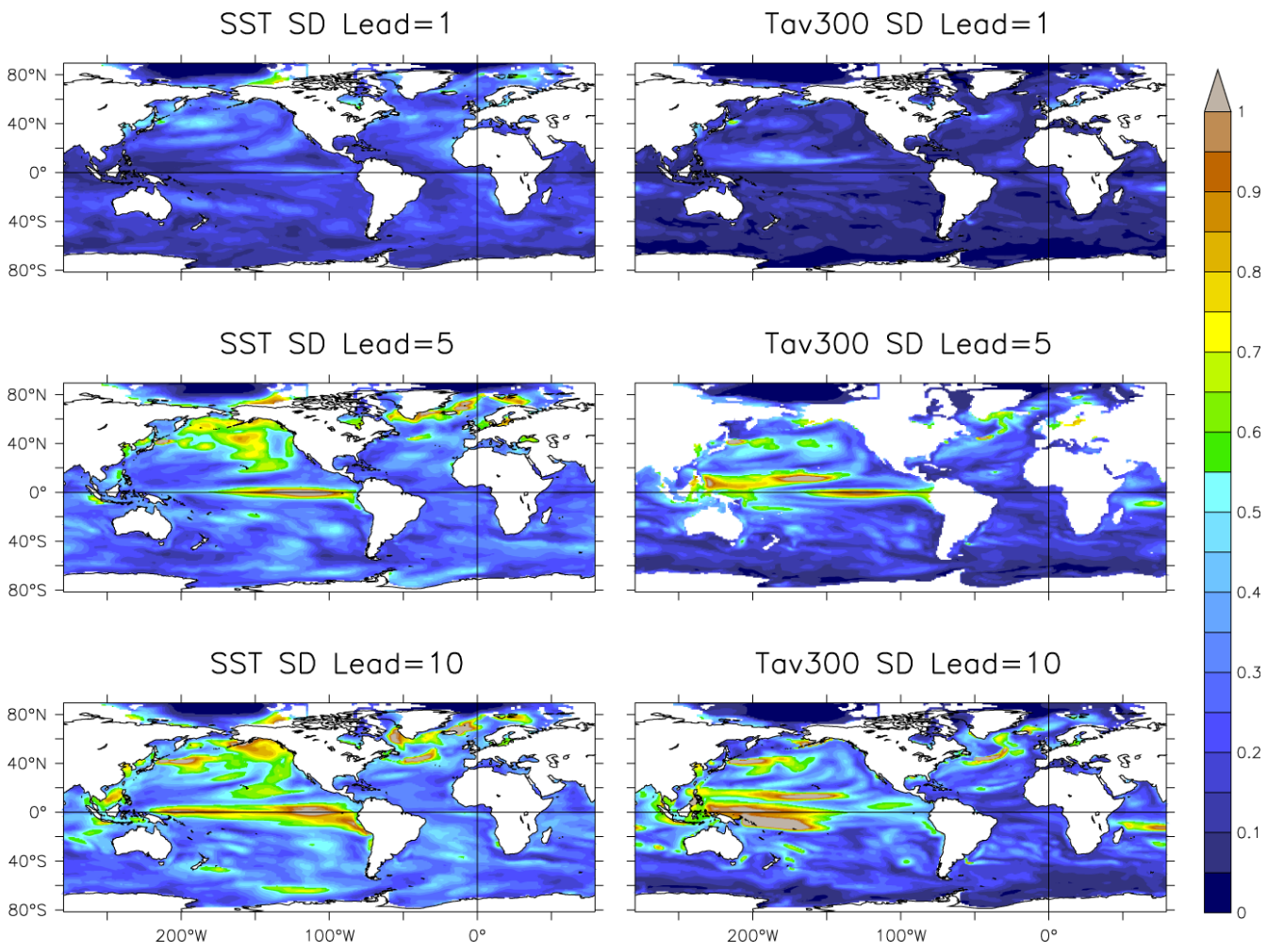
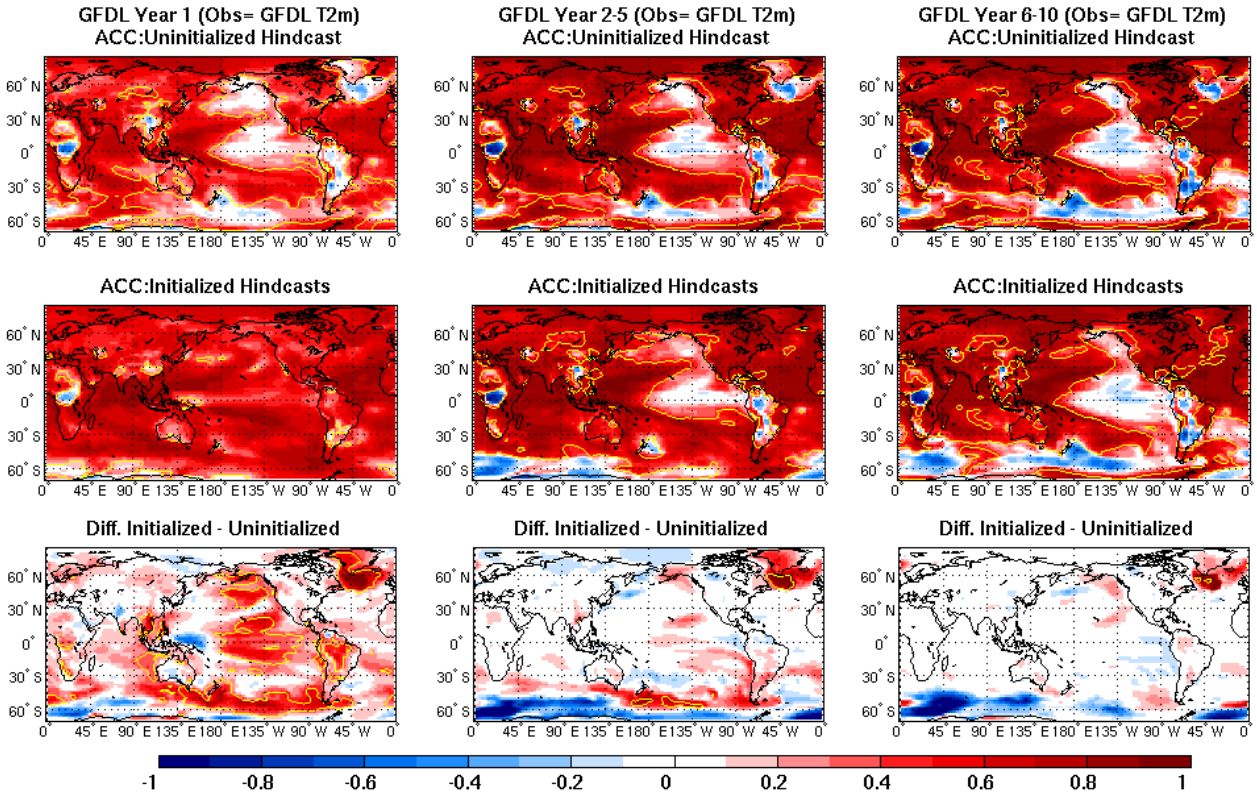


Figure 2 – Ensemble Standard Deviation of SST (left) and average ocean temperature over top 300 meters, Tav300 (right) from hindcasts for leads 1,5,10. Units are degrees C.

601



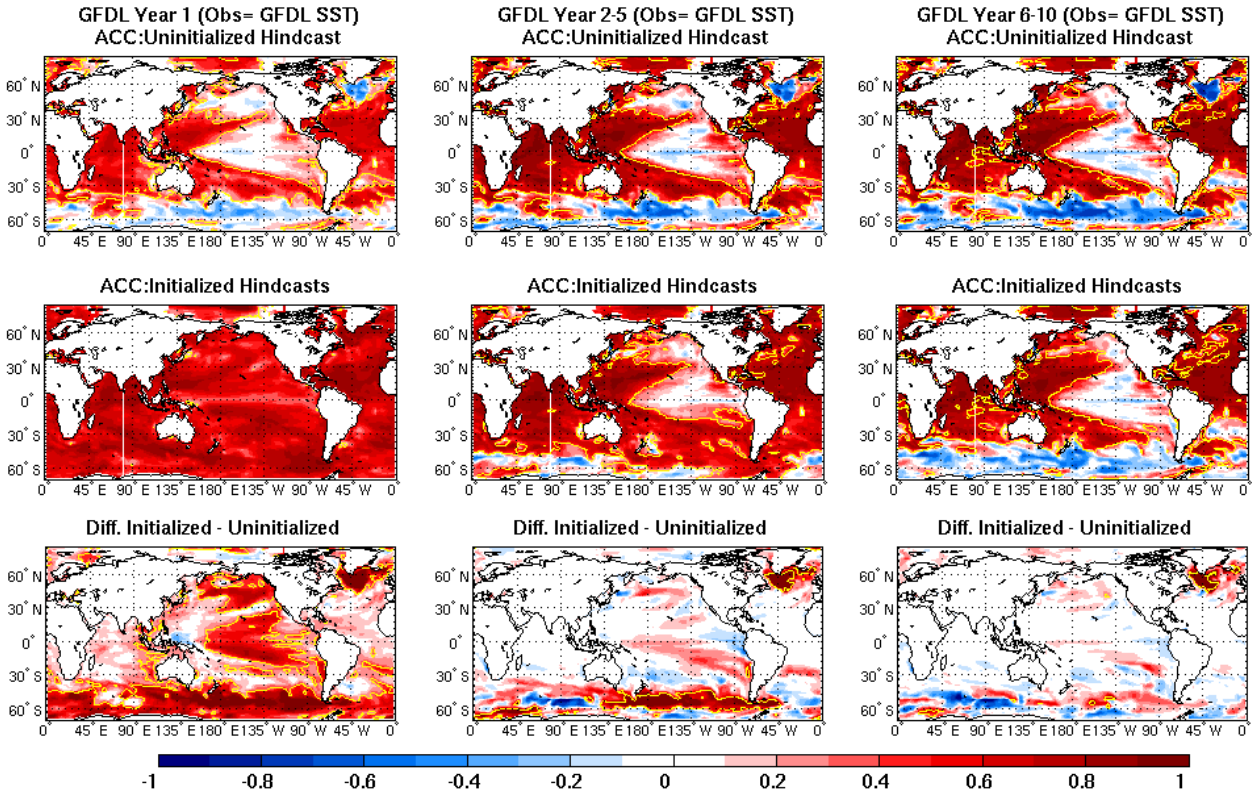
602

603

604 Figure 3 – T2m Anomaly Correlation Coefficient (ACC) with the GFDL reanalysis of the uninitialized
605 forecasts (top row), initialized forecasts (middle row), and their difference (bottom row). Forecast year
606 1 is the first column, the average of forecast years 2-5 is the second column and the average of forecast
607 years 6-10 the third column. The yellow contour lines represent the 90% significance.

608

609



610

611

612 Figure 4 – Same as in Fig 3 but for SST.

613

614

615

616

617

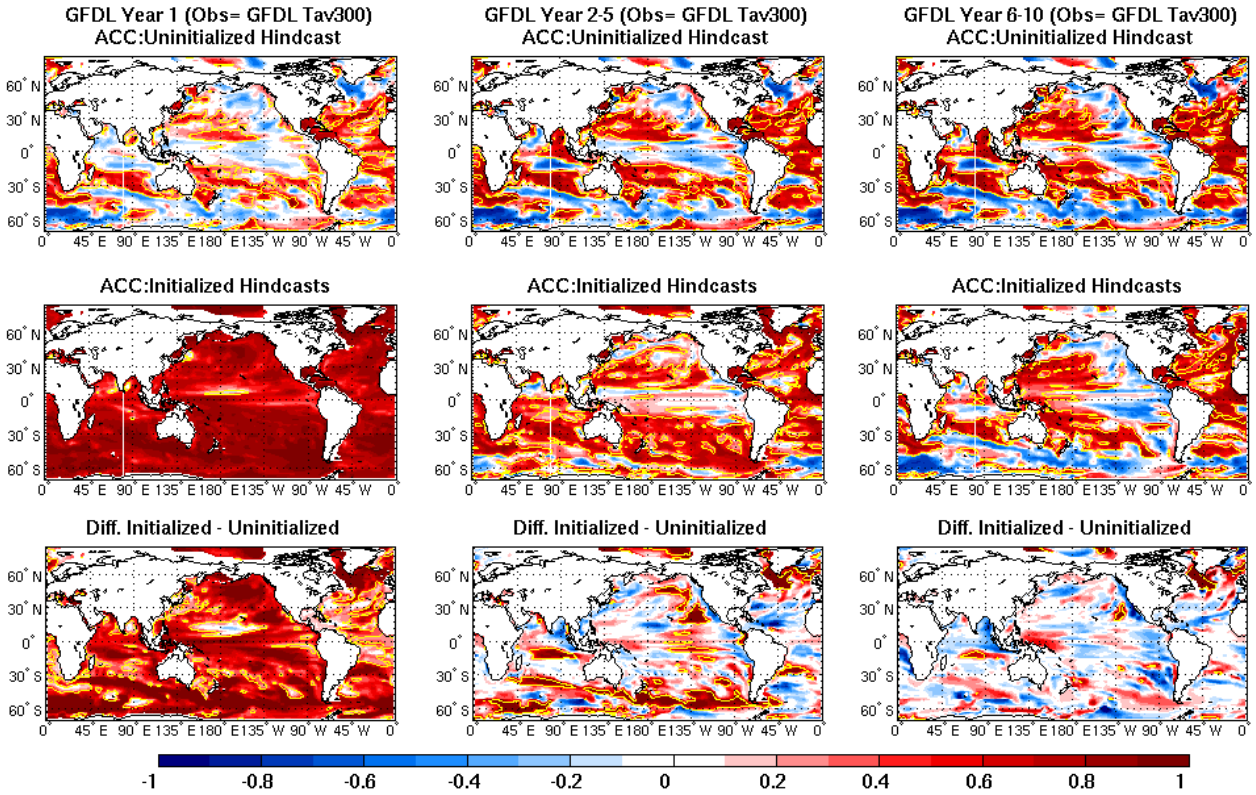
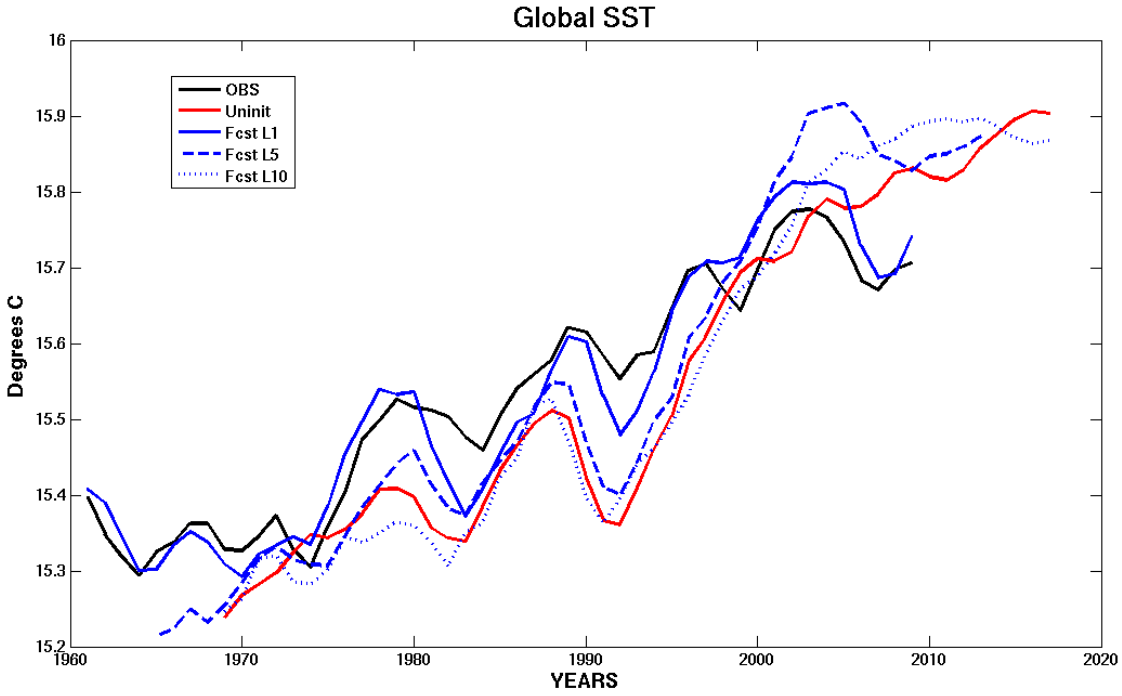
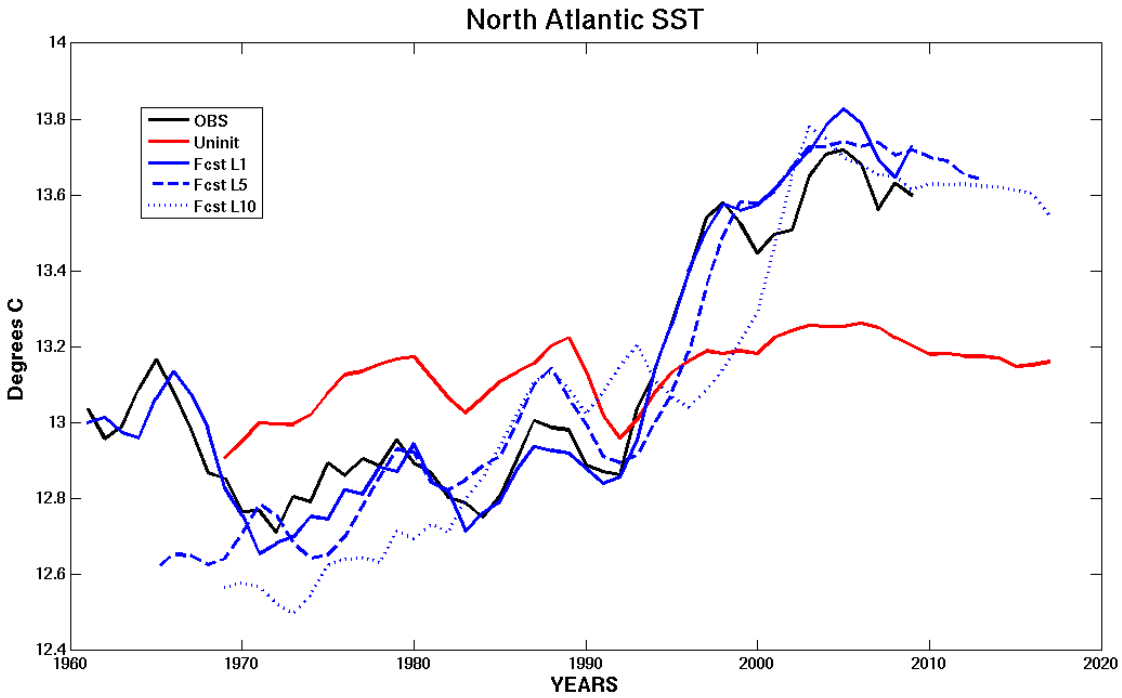


Figure 5 – Same as in Fig. 3 but for Tav300.



631

632



633

634

635 Figure 6 – Two year running mean bias corrected average SST, black - OBS observations from

636 ECDAv3.1, red – Uninitialized , blue solid lead 1 forecast, blue dash lead 5, blue dot lead10.

637 Top - Global SST, Bottom – North Atlantic SST averaged over 30N-65N, 80w-10w. Units are

638 degrees C.

639

640

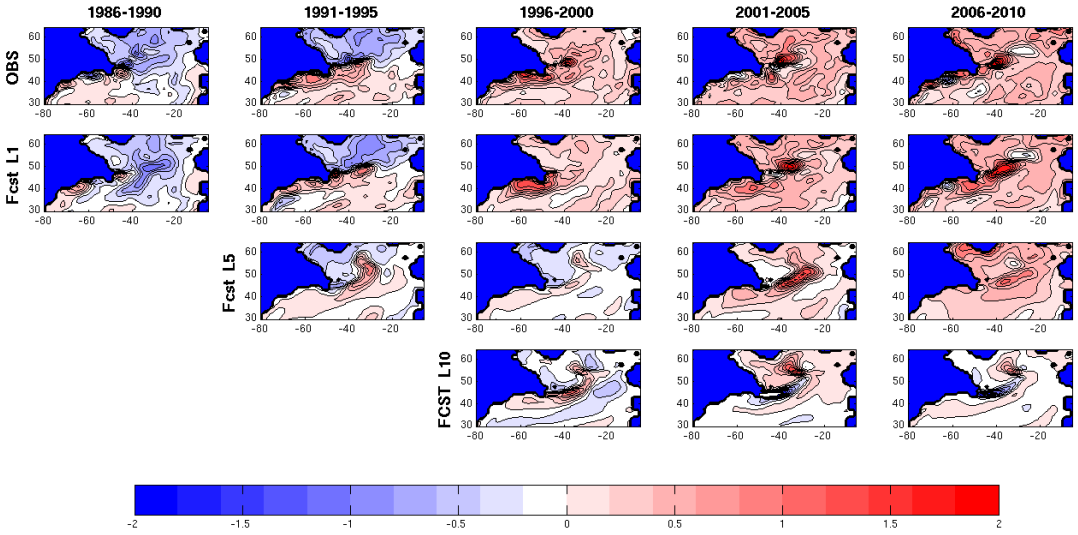


Figure 7 – North Atlantic Tav300 pentads in degrees C. Top row shows observations. Subsequent rows show forecast leads 1, 5, and 10. Leads are in years.

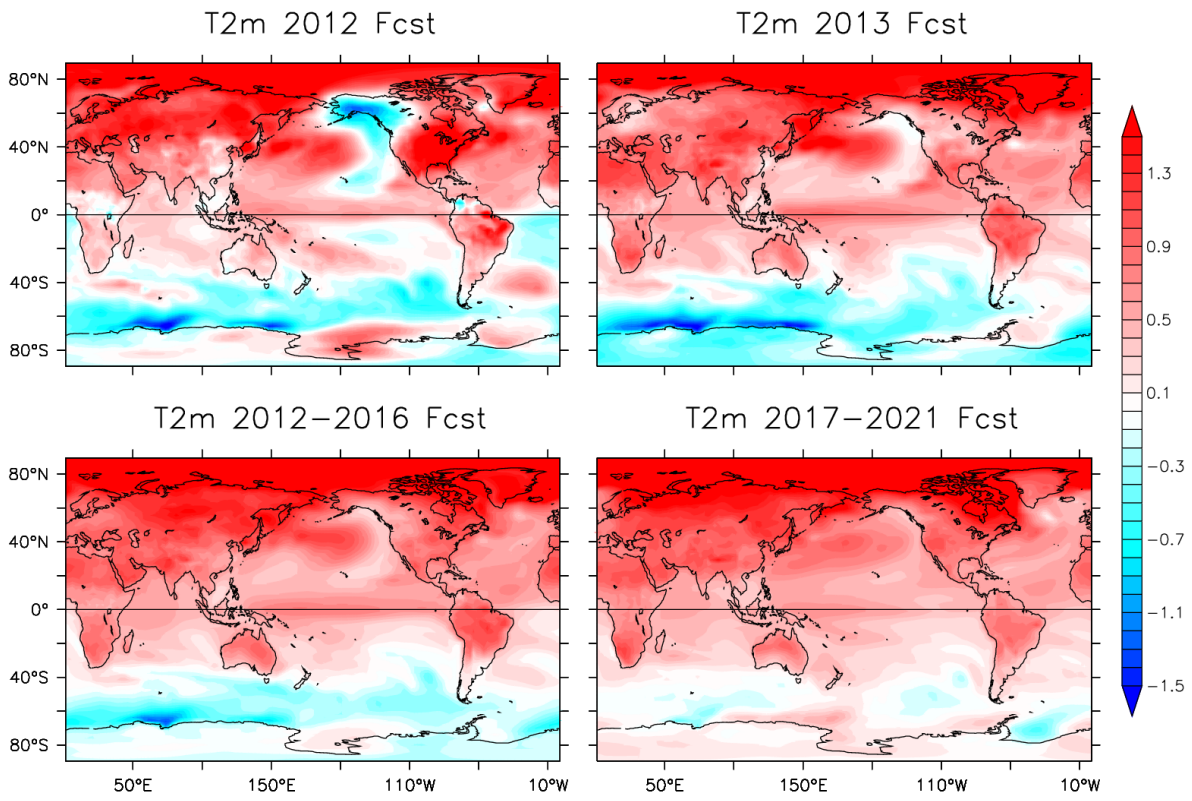
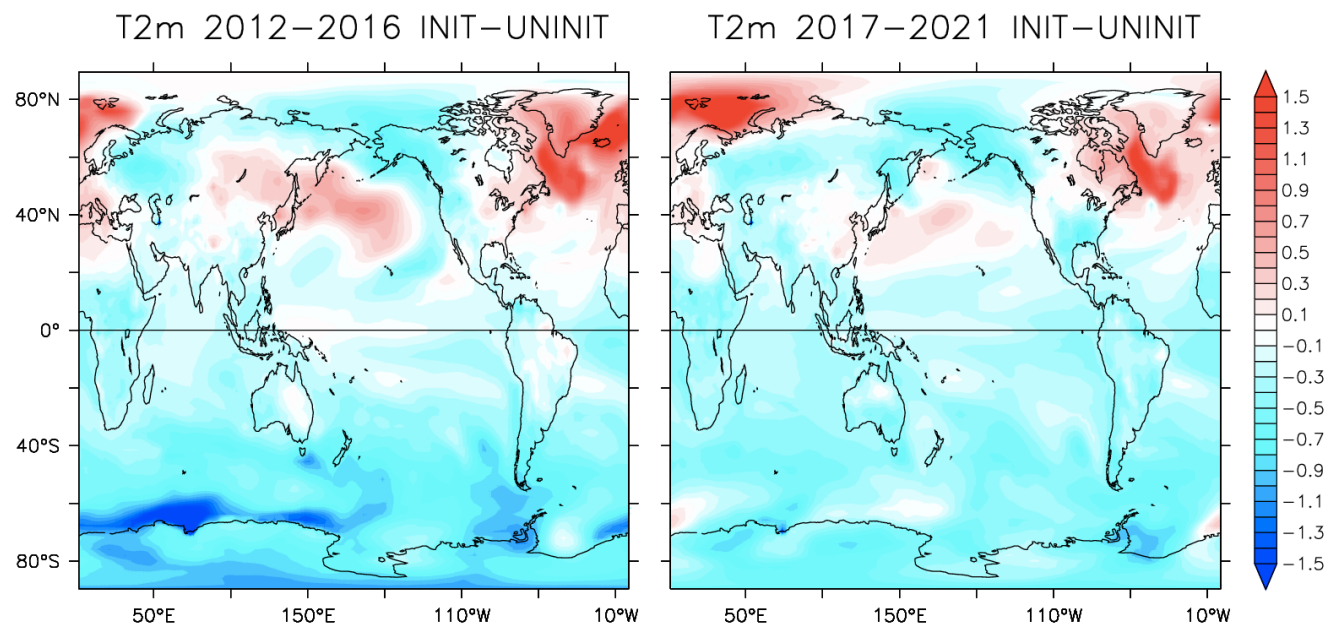


Figure 8 – T2m forecast bias corrected anomaly relative the to 1971-2000 climatology. Units are in degrees C.



655

656

657 Figure 9 – Difference of initialized forecast (start date is January 1, 2012) and unitialized projection,

658 showing impact of initialization. Left – 2012-2016 forecast period, Right – 2016-2021 forecast period.

659 Contour units are in degrees C.

660

661